Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

SOLUTION OF THE MATRIX EQUATION $X^{-1}AX = N$.

BY H. T. BURGESS.

If A is any square matrix, it is a well-known theorem in the matrix theory that there exists a similar matrix N which is of a unique canonical form.* The form of N is determined by the elementary divisors of the characteristic matrix $\lambda I + A$. The existence of the matrix X is known, but no simple method of computing it is to be found in the literature. It is the object of this paper to give a simple and direct method.

In the *Annals of Mathematics*, Vol. XVIII, No. 1, September, 1916, p. 4, is found a brief paper to which reference will be made as the former paper.

1. **Notation.** Consider the linear substitution,

$$\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 + \cdots + b_{1n}x_n, \\ y_2 &= b_{21}x_1 + b_{22}x_2 + \cdots + b_{2n}x_n, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ y_n &= b_{n1}x_1 + b_{n2}x_2 + \cdots + b_{nn}x_n, \end{aligned}$$

and observe that it may be written in matrix notation in the form

$$\left\|\begin{matrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{matrix}\right\| = \left\|\begin{matrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{matrix}\right\| \cdot \left\|\begin{matrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{matrix}\right\|,$$

in which the ordinary rule of matrix multiplication applies. More briefly we may use $y = Bx$ in which small letters denote matrices of n rows and one column, and the large letter denotes a square matrix of order n. If the matrix y is zero, we have a system of n linear homogeneous equations in n unknowns expressed in this notation by $Bx = 0$.

It immediately appears from this notation that the system of linear homogeneous equations $DBx = 0$ has each and every solution of $Bx = 0$ as a solution. Further if the rank of the matrix DB is less than that of the matrix B, the system $DBx = 0$ has solutions which are not solutions of the system $Bx = 0$. If κ is any positive integer, every solution of $Bx = 0$ is also a solution of $B^k x = 0$, while any solution of $B^k x = 0$ which is not a solution of $B^{k-1}x = 0$ is not a solution of $B^{\sigma}x = 0$ for $\sigma < \kappa$. This of course applies only when B is a singular matrix.

---

* For a demonstration of this theorem, see Hilton, Linear Substitutions, p. 29.

2. **Method of computation.** By Theorem 2 in the former paper, we can compute the elementary divisors of the characteristic matrix $\lambda I + A$, and thus uniquely determine the form of the canonical matrix N for any given matrix A. The nature of the canonical form N is found in the "fundamental theorem" in the former paper and need not be restated here.

From (a) $X^{-1}AX = N$, we have (b) $AX = XN$. Any solution X of (a) is also a solution of (b), and any non-singular solution of (b) is also a solution of (a). If we can devise a scheme of finding all of the solutions of (b), we have only to select a non-singular solution of (b) to obtain a solution of (a).

The necessary and sufficient conditions for X to be a solution of (b) are found by equating the corresponding elements by columns in the products AX and XN. Referring to the "fundamental theorem" in the former paper, we see that it is sufficient to consider only the t_ϵ columns corresponding to the typical component N_{t_ϵ} in N. For this purpose, we write (b) in the form $(\alpha_\epsilon I + A)X = X(\alpha_\epsilon I + N)$. It is further obvious that the conditions for these t_ϵ columns can be found by considering only the e_π columns corresponding to the typical component E_{e_π} in N_{t_ϵ}. Using the notation previously explained, we denote the block of columns in X which corresponds to these e_π columns by

$$x_i^{(\pi)} = \left\| \begin{array}{c} x_{1i}^{(\pi)} \\ x_{2i}^{(\pi)} \\ \cdots \\ x_{ni}^{(\pi)} \end{array} \right\| \qquad (i = 1, 2, \cdots, e_\pi).$$

Since in any column of the product $(\alpha_\epsilon I + A)X$, only elements from the corresponding column of X appear, we get $(\alpha_\epsilon I + A)x_{i+1}^{(\pi)}$ in the column corresponding to the $(i + 1)$st column of the component E_{e_π}. Likewise for $i > 0$, in the corresponding column of the product $X(\alpha_\epsilon I + N)$, we get simply $x_i^{(\pi)}$; for the only non-zero element in the $(i + 1)$st column of $\alpha_\epsilon I + E_{e_\pi}$ is a 1 appearing in the over-principal diagonal. For $i = 0$, we get $x_0^{(\pi)} = 0$, for the first column of $\alpha_\epsilon I + E_{e_\pi}$ contains only zeros. We can now express the τ sets of equations to determine the τ blocks of columns in X which correspond to the t_ϵ columns in the component N_{t_ϵ} as follows:

$$(c) \quad \begin{array}{ll} x_i^{(\pi)} = (\alpha_\epsilon I + A)x_{i+1}^{(\pi)}, & i = 0, 1, 2, \cdots e_\pi, \\ x_i^{(\pi)} = 0, \quad i = 0; \quad x_i^{(\pi)} \neq 0, \quad i \neq 0, & \pi = 1, 2, \cdots \tau. \end{array}$$

For convenience denote $\alpha_\epsilon I + A$ by B, and consider the following set of matrix equations:

$$
(d)\qquad
\begin{array}{r}
x_0^{(\pi)} = Bx_1^{(\pi)} = B^2x_2^{(\pi)} = B^3x_3^{(\pi)} = \cdots = B^{\kappa-1}x_{\kappa-1}^{(\pi)} = B^{\kappa}x_{\kappa}^{(\pi)} \\
x_1^{(\pi)} = Bx_2^{(\pi)} = B^2x_3^{(\pi)} = \cdots = B^{\kappa-2}x_{\kappa-1}^{(\pi)} = B^{\kappa-1}x_{\kappa}^{(\pi)} \\
x_2^{(\pi)} = Bx_3^{(\pi)} = \cdots = B^{\kappa-3}x_{\kappa-1}^{(\pi)} = B^{\kappa-2}x_{\kappa}^{(\pi)} \\
x_3^{(\pi)} = \cdots = B^{\kappa-4}x_{\kappa-1}^{(\pi)} = B^{\kappa-3}x_{\kappa}^{(\pi)} \\
\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\
x_{\kappa-1}^{(\pi)} = Bx_{\kappa}^{(\pi)}
\end{array}
$$

From this notation, it is obvious that this set of matrix equations is derivable from those equations which constitute the diagonal by direct matrix multiplication. If we choose $x_\kappa^{(\pi)}$ as a solution of $B^\kappa x = 0$ and not a solution of $B^{\kappa-1}x = 0$, we have $x_0^{(\pi)} = 0$, and $x_1^{(\pi)}$, $x_2^{(\pi)}$, $\cdots$ $x_\kappa^{(\pi)} \neq 0$. To obtain a solution of the sets of equations (c), we take simultaneously

$$
\begin{array}{l}
\pi = 1, \quad 2, \quad \cdots \quad \tau, \\
\kappa = e_1, \quad e_2, \quad \cdots \quad e_\tau.
\end{array}
$$

This choice is evidently always possible, for the rank of B^κ is less than the rank of $B^{\kappa-1}$ for $\kappa \leqq e_1$ by Theorem 2 of the former paper. Moreover the choice of $x_\kappa^{(\pi)}$ is by no means unique, there being an infinity of ways of making the selection.

Applying this same process to the remaining roots of the characteristic equation of A, we obtain the complete matrix X which is a solution of equation (b).

Since all the solutions of (b) are to be obtained in this manner, to obtain a non-singular solution, it is only necessary to properly choose the matrices whose selection is by no means unique.

In the application of this method, we can greatly simplify the actual computation of the elementary divisors and the matrix X if we observe first, that the reduction of any matrix B by elementary transformations on its rows only is equivalent to a matrix multiplication of the type TB where T is non-singular; secondly, that B and TB have the same rank; and thirdly, that the two systems of linear homogeneous equations $Bx = 0$ and $TBx = 0$ are equivalent. Consider the matrix equations

(1) $T_1B = R_1$ where r_1 is the rank of B and R_1,

(2) $T_2T_1B^2 = T_2R_1B = R_2$ where r_2 is the rank of B^2 and R_2,

. . . .

$(\kappa)\quad T_\kappa \cdots T_2T_1B^\kappa = T_\kappa R_{\kappa-1}B = R_\kappa$ where r_κ is the rank of B^κ and R_κ,

. . . .

Let T_κ be a non-singular matrix such that $T_\kappa R_{\kappa-1}B = R_\kappa$ where R_κ is reduced by elementary transformations on its rows only, then the ranks of B^κ and R_κ are the same and the two systems of linear homogeneous equations $B^\kappa x = 0$ and $R_\kappa x = 0$ are equivalent. Furthermore by the former paper, Theorem 2, we need to carry the work out only until we find $n - r_\kappa = t_\epsilon$, where t_ϵ is the multiplicity of the factor $\lambda - \alpha_\epsilon$ in the characteristic equation of the matrix A.

3. **Illustration.** Given the matrix

$$A = \begin{Vmatrix} -1 & 6 & 1 & 2 & 4 & 2 & 0 & 2 & 4 & 0 \\ 1 & 2 & 0 & 14 & 6 & 6 & -6 & 0 & 8 & -2 \\ 8 & 14 & -1 & 4 & 6 & 4 & 2 & 8 & 0 & 2 \\ 4 & 7 & 3 & -8 & 3 & -2 & 5 & 4 & -2 & 3 \\ -3 & -9 & -3 & -8 & -11 & -2 & 2 & -2 & -3 & -1 \\ 6 & 7 & 1 & -2 & 1 & 2 & 4 & 6 & 3 & 1 \\ -8 & -8 & 0 & 0 & 0 & -8 & -6 & -8 & -8 & 0 \\ 6 & 2 & 3 & -24 & -6 & -8 & 14 & 4 & -16 & 6 \\ -2 & -3 & -1 & 2 & -1 & 0 & -2 & -2 & -1 & -1 \\ 0 & -6 & -1 & -4 & -6 & 4 & 2 & 0 & 16 & -4 \end{Vmatrix}.$$

Its characteristic equation is found to be $\varphi(\lambda) = (\lambda - 4)(\lambda - 3)^2(\lambda - 2)^7 = 0$, and this shows that N is of the form

$$N = \begin{Vmatrix} N_1 & & \\ & N_2 & \\ & & N_7 \end{Vmatrix}.$$

For the first column of X, we have to solve the matrix equation

$$(4I + A)x = 0.$$

The solution found in the usual way is $x_1 = (0, 1, 0, 0, -1, 0, 0, -1, 0, 0)$.

For the second and third columns of X, we must compute, from $B = 3I + A$,

$$R_1 = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{Vmatrix}, \qquad r_1 = 9;$$

$$R_2 = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{Vmatrix}, \qquad r_2 = 8.$$

By Theorem 2 of the former paper, we have

$$N_2 = \begin{Vmatrix} -3 & 1 \\ 0 & -3 \end{Vmatrix},$$

and our equations to determine x_2 and x_3 are

$$0 = (3I + A)x_2 = (3I + A)^2 x_3, \tag{1}$$

$$x_2 = (3I + A)x_3. \tag{2}$$

We take x_3 as a solution of $R_2 x = 0$ and not a solution of $R_1 x = 0$. Such a solution is obviously $x_3 = (0, 1, 1, 0, -1, 0, 0, -1, 0, -1)$, and from (2) we compute $x_2 = (1, 1, 0, 0, -1, 0, 0, -2, 0, 0)$.

For the last seven columns of X, we compute from $B = 2I + A$

$$R_1 = \left\| \begin{matrix} 10 & 0 & 0 & 0 & 0 & -22 & 1 & 1 & 0 & 12 \\ 0 & 10 & 0 & 0 & 0 & 32 & 4 & 9 & 0 & -12 \\ 0 & 0 & 10 & 0 & 0 & 64 & 8 & 8 & 0 & -24 \\ 0 & 0 & 0 & 4 & 0 & 10 & -1 & 1 & 0 & -4 \\ 0 & 0 & 0 & 0 & 10 & -66 & -7 & -12 & 0 & 26 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{matrix} \right\|, \qquad r_1 = 6;$$

$$R_2 = \left\| \begin{matrix} 81 & 0 & 0 & 0 & -47 & 132 & 41 & 64.5 & 262 & -25 \\ 0 & 972 & 0 & 0 & 584 & -744 & -20 & 174 & -2380 & 352 \\ 0 & 0 & 486 & 0 & 584 & -744 & -20 & -312 & -2380 & 352 \\ 0 & 0 & 0 & -972 & -410 & 276 & 530 & 249 & 2320 & -94 \end{matrix} \right\|,$$

$$r_2 = 4;$$

$$R_3 = \left\| \begin{matrix} 360 & 0 & -233 & 40 & -472 & 932 & 170 & 426 & 2210 & -276 \\ 0 & 2 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 5177 & -3160 & 4888 & -7028 & 1510 & -2514 & -17810 & 3444 \end{matrix} \right\|,$$

$$r_2 = 3.$$

We have, from these ranks,

$$N_7 = \left\| \begin{matrix} E_3 & & & \\ & E_2 & & \\ & & E_1 & \\ & & & E_1 \end{matrix} \right\|,$$

and the following equations to determine the last seven columns of X:

$$0 = (2I + A)x_4 = (2I + A)^2 x_5 = (2I + A)^3 x_6, \tag{1}$$

$$x_4 = (2I + A)\, x_5 = (2I + A)^2 x_6, \tag{2}$$

$$x_5 = (2I + A)\, x_6, \tag{3}$$

$$0 = (2I + A)x_7 = (2I + A)^2 x_8, \tag{4}$$

$$x_7 = (2I + A)\, x_8, \tag{5}$$

$$0 = (2I + A)x_9, \tag{6}$$

$$0 = (2I + A)x_{10}. \tag{7}$$

We must choose x_6 as a solution of $R_3 x = 0$ and not a solution of $R_2 x = 0$. Such a solution is $x_6 = (390, 492, 102, 0, -492, 1, 0, -882,$

-1, -102). Equations (3) and (2) give

$$x_5 = (-290, -392, -98, 0, 391, 1, 0, 686, -1, 90),$$

$$x_4 = (194, 306, 112, -11, -288, 16, -32, -500, 0, -88).$$

Next we choose x_8 as a solution of $R_2x = 0$ and not a solution of $R_1x = 0$. Such a solution is

$$x_8 = (-34, -45, -10, 0, 45, 0, -1, 80, 0, 10),$$

and equation (5) gives

$$x_7 = (26, 42, 16, -1, -40, 2, -4, -68, 0, -12).$$

It remains to choose x_9 and x_{10}, as solutions of $R_1x = 0$, care being taken that they are independent of x_4 and x_7. They may be taken as

$$\begin{aligned} x_9 &= (-2, -8, -16, 5, 14, 0, 20, 0, 0, 0), \\ x_{10} &= (2, 18, 16, 5, -24, 0, 0, -20, 0, 0). \end{aligned}$$

This gives

$$X = \begin{Vmatrix} 0 & 1 & 0 & 194 & -290 & 390 & 26 & -34 & -2 & 2 \\ 1 & 1 & 1 & 306 & -392 & 492 & 42 & -45 & -8 & 18 \\ 0 & 0 & 1 & 112 & -98 & 102 & 16 & -10 & -16 & 16 \\ 0 & 0 & 0 & -11 & 0 & 0 & -1 & 0 & 5 & 5 \\ -1 & -1 & -1 & -288 & 391 & -492 & -40 & 45 & 14 & -24 \\ 0 & 0 & 0 & 16 & 1 & 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -32 & 0 & 0 & -4 & -1 & 20 & 0 \\ -1 & -2 & -1 & -500 & 686 & -882 & -68 & 80 & 0 & -20 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -88 & 90 & -102 & -12 & 10 & 0 & 0 \end{Vmatrix}.$$

University of Wisconsin,
Madison, Wisconsin.